AFHRL-TR-90-90

AD-A234 292

AIR FORCE RESEARCH TO LINK STANDARDS FOR ENLISTMENT TO ON-THE-JOB PERFORMANCE

Mark S. Teachout
Martin W. Pellum, Major, USAF

TRAINING SYSTEMS DIVISION
Brooks Air Force Base, Texas 78235-5601

February 1991

Interim Technical Report for Period October 1984 – December 1990

HUMAN RESOURCES

LABORATORY

91 4 04 030

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235-5601

NOTICE

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

The Public Affairs Office has reviewed this report, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This report has been reviewed and is approved for publication.

HENDRICK W. RUCK, Technical Advisor
Training Systems Division

HAROLD G. JENSEN, Colonel, USAF
Commander

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	February 1991	Interim Report – October 1984 – December 1990

4. TITLE AND SUBTITLE
Air Force Research to Link Standards for Enlistment to On-the-job Performance

5. FUNDING NUMBERS
C – F41689-86-D-0052
PE – 63227F
PR – 2922
TA – 01
WU – 01

6. AUTHOR(S)
Mark S. Teachout
Martin W. Pellum

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Training Systems Division
Air Force Human Resources Laboratory
Brooks Air Force Base, Texas 78235-5601

8. PERFORMING ORGANIZATION REPORT NUMBER
AFHRL-TR-90-90

9. SPONSORING/MONITORING AGENCY NAMES(S) AND ADDRESS(ES)

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
This report appeared as Chapter 6 of "A Report to the House Committee on Appropriations: Sixth Annual Report to the Congress on Joint-Service Efforts to Link Standards for Enlistment to On-the-job Performance," December, 1987.

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

This report describes Air Force research supporting the Joint-Service Job Performance Measurement/ Enlistment Standards Project. This project has focused on developing a technology for assessing the performance capability of enlisted personnel, with the goal of determining relationships among selection and classification standards and on-the-job performance. This report describes those relationships. A brief overview of the Air Force Job Performance Measurement (JPM) research program is followed by results of analyses conducted for eight Air Force enlisted specialties of relationships among enlistment test scores, job experience, educational attainment, and hands-on job performance test scores. Additional analyses indicating the reliability and validity of the JPM technology are presented for four Air Force specialties. Finally, plans for applying the JPM technology to selection and classification systems and training systems are described.

14. SUBJECT TERMS			15. NUMBER OF PAGES
Armed Forces Qualification Test	job experience	validity	18
educational attainment	job performance measurement	Walk-Through Performance Test	16. PRICE CODE
hands-on performance tests	reliability		

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UI

SUMMARY

The Air Force Human Resources Laboratory (AFHRL) has participated in a Joint-Service Job Performance Measurement Project since 1980. The primary purpose of this project was to validate selection and classification tests (i.e., the Armed Services Vocational Aptitude Battery) against hands-on measures of job performance. This report documents Air Force research and development (R&D) activities to validate the ASVAB for eight Air Force enlisted specialties. Additional Air Force R&D is described that assesses the reliability and validity of the Air Force's Job Performance Measurement System (JPMS). Finally, the report describes plans for using job performance information to improve the Air Force's selection and classification and training systems.

PREFACE

This document appeared as Chapter 6 in the 1987 Report to the House Appropriations Committee on Service efforts to link enlistment standards to job performance. The authors wish to thank Drs. William E. Alley and Hendrick W. Ruck, whose guidance and comments were helpful throughout this effort; Ms Doris Black and the programming support staff, who assisted in most of the data analyses; Drs. Kurt Kraiger and Terry Dickinson, who contributed to the reliability and validity analyses and whose work is documented more fully in other technical papers and reports; Lt Col Nick Ovalle for his wise advice; Col Rodger D. Ballentine, who led the JPM project through its formative years; and especially to Drs. Jerry W. Hedge and M. Suzanne Lipscomb for their years of hard work that contributed to the success of this project. Major Pellum is now at the Directorate of Commissioning Programs, Headquarters, Air Training Command.

TABLE OF CONTENTS

LIST OF TABLES

AIR FORCE RESEARCH TO LINK STANDARDS FOR ENLISTMENT TO ON-THE-JOB PERFORMANCE

I. INTRODUCTION

Air Force research supporting the Joint-Service Job Performance Measurement/Enlistment Standards Project has focused on developing a technology for assessing the performance capability of enlisted personnel, with the goal of determining the relationship between Air Force selection and classification standards and on-the-job performance. Eight Air Force specialties (AFSs), or career fields, were selected for developing a prototype Job Performance Measurement System (JPMS). These were selected using several criteria; for example, the number of first-term airmen assigned to the specialty, the Armed Services Vocational Aptitude Battery (ASVAB) composites (specific combinations of subtests) used by the Air Force for classifying recruits into the specialty, and the similarity of the specialty to jobs in the other Services in a way that could facilitate cross-Service technology transfer. The eight AFSs under study are listed below.

ASVAB Composite	AFS	Description
Mechanical	426X2	Jet Engine Mechanic
Administrative	492X1	Information Systems Radio Operator
General	272X0	Air Traffic Control Operator
Electronics	328X0	Avionic Communications Specialist
Mechanical	423X5	Aerospace Ground Equipment Specialist
Administrative	732X0	Personnel Specialist
General	122X0	Aircrew Life Support Specialist
Electronics	324X0	Precision Measuring Equipment Specialist

This report begins with a brief overview of the Air Force Job Performance Measurement (JPM) research and development (R&D) program and then presents the general results of data collected in the above eight specialties. In-depth, Air Force-specific analyses are reported for the first set of four specialties only. The report concludes by noting R&D currently planned or underway with regard to the JPM technology.

II. THE AIR FORCE JPM TECHNOLOGY

The technology employed in the Air Force Job Performance Measurement Program centers around the development and administration of several types of measurement instruments. These procedures have been described in detail in previous reports (Hedge & Teachout, 1986; Lipscomb & Hedge, 1988) and so are described only briefly here.

The Measures

Work sample tests have been consistently identified as the highest fidelity measures of job performance capability. In most cases, work sample tests employ hands-on performance measures which require incumbents to display the same behaviors as they would on the job (i.e., perform the tasks using operational equipment, materials, and procedures. As with the other Services, the Air Force developed hands-on measures for each AFS under study. In addition, other measurement techniques are being explored as feasible alternatives to hands-on testing where the latter may not be practical. Performance interviews are one method that is

unique to the Air Force JPMS. As with hands-on testing, the interviews take place at the job site; however, instead of actually performing the task, the incumbent describes, in a show-and-tell fashion, the procedures he/she would follow if performing the task. The combination of hands-on and interview testing methods is referred to as Walk-Through Performance Testing (WTPT). Performance interviews were developed for all eight AFSs.

A series of rating forms are also included in the Air Force JPMS. Four types of forms have been developed. These include task-level ratings, dimension-level ratings (identified via cluster analyses of the tasks), and global ratings (single-item scales for task proficiency and interpersonal proficiency). The fourth form is referred to as the Air Force-wide rating form and includes dimension-level scales of factors deemed important to successful performance in the Air Force, such as leadership, initiative/effort, and self-development. Rating forms were also developed for all eight AFSs.

Written job knowledge tests (JKTs) are the last measurement method encompassed in the Air Force JPMS. These were constructed for the last four AFSs only. Three of the JKTs were developed using Army procedures in a cooperative effort involving cross-Service transfer of technology. The fourth was constructed following traditional Air Force JKT development procedures.

JPMS Development

Development of JPMS measures begins with the selection of tasks which represent the jobs of first-term airmen within the specialty. Two primary sources of information are used to define this job domain: job inventory data collected by the Air Force Occupational Measurement Center, and documents outlining technical training programs for the specialty. After being categorized into four groups based on difficulty, tasks are randomly selected from each difficulty quartile following the procedures specified in Lipscomb and Hedge (1988). The resulting task lists are then presented to subject-matter experts (SMEs), who judge each task in terms of its representativeness of the job domain and its amenability to performance testing.

A field-intensive task analysis procedure then begins which breaks each task down into its subcomponents, or steps, and identifies the associated equipment, tools, and procedures required to perform the task. The number of tasks that resulted in hands-on test items in the eight AFSs studied ranged from 10 for AFS 426X2 to 21 for AFS 324X0. Between 6 and 15 additional tasks were tested via the interview method, with approximately one-third of these overlapping with tasks included in the hands-on test. Similarly, the rating scales included all the hands-on tasks plus numerous others selected from the job domain. Job knowledge tests contained between 100 and 301 items that corresponded to those tasks associated with the WTPT.

Data Collection

Training of test administrators is an essential first step in the data collection process. The test administrator training program employed by the Air Force incorporates all key elements identified in the professional literature for obtaining both accurate and reliable performance information. The content of the training program emphasizes proper administration of hands-on and interview tests, from setting up equipment to scoring each step in each task. Several different training methods are used. These include role-playing exercises and scoring of videotapes depicting correct and incorrect task performances. In addition, a technique referred to as "shadow scoring" is used both in scoring the training videotapes and during data collection in the field. In shadow scoring, raters independently observe and score an individual performing

a task. The raters then compare and discuss their ratings using standard scoring criteria. This technique has been shown to be extremely effective in increasing interrater agreement in the scoring process. The training program proved to be successful for both civilian contractor test administrators employed for the 426X2 AFS and active duty SMEs serving as administrators for the other seven specialties.

All data were collected in the operational (field) environment using actual equipment and materials. The majority of test incumbents were active duty airmen in their first term of enlistment (less than 4 years in the Service) who had a minimum of 3 months of job experience. For the eight specialties studied, a total of 1,493 airmen were tested across 70 different Air Force bases. Performance rating forms were completed by each incumbent and his/her supervisor. In addition, over 3,400 rating forms were completed by peers (coworkers) of the incumbents.

A considerable amount of additional data is collected at the time the incumbent is tested. A primary piece of information is the frequency and recency of the incumbent's experience on the task being evaluated. Incumbents also complete a questionnaire at the conclusion of the testing session which addresses various factors related to performance, such as work motivation, job satisfaction, situational constraints, and acceptability of the JPMS. Additionally, technical training school grades for each incumbent are collected along with data from their personnel files, in particular ASVAB scores and education level. Hands-on, interview, job knowledge, and rating form scores, as well as training data and information obtained from personnel files, make up the Air Force JPMS data base.

III. RESULTS OF DATA ANALYSES

Common Data Analyses

Analyses reported in this section are based on data collected for the eight AFSs. These analyses were directed by the Principal Deputy Assistant Secretary of Defense for Force Management and Personnel to maintain consistency across the Services in reporting results.

Sample Description. All individuals in the sample held 3-level (apprentice) or 5-level (journeyman) skill ratings, and most were in their first term of enlistment. Each sample within an AFS approximated its respective first-term population with regard to race, gender, and aptitude. Table 1 presents descriptive statistics for four variables: hands-on performance test (HOPT) scores, job experience, aptitude, and educational attainment. HOPT scores have been converted to standard T-scores to facilitate comparison across specialties. The average experience level ranged from 23 to 35 months in Service, and the average aptitude ranged from the 56th to the 80th percentile on the Armed Forces Qualification Test (AFQT). Most of the airmen were high school diploma graduates at the time they enlisted in the Air Force.

Table 2 presents standardized HOPT scores for each sample, broken into four subgroups for experience and two subgroups for aptitude. This information should be interpreted with caution due to the small sample sizes in some of the subgroups, especially in making comparisons across groups.

Reliability. Three indices of reliability are presented in Table 3. Scorer reliability represents the interrater agreement calculated using the shadow scoring of the hands-on measures in the field. Clearly, the high reliabilities indicate the integrity of the test administrator training program and the great care taken in maintaining quality throughout the data collection process.

Table 1. **Descriptive Statistics for HOPT, Job Experience, Aptitude, and Educational Attainment[a]**

AFS		HOPT[b]	Job Experience[c]	Aptitude[d]	Educational Attainment[e] (N)	
	Mean	50	29.3	59.0	186	HSDG
122X0	SD	10	11.0	17.3	2	NHSDG
	N	195	195	172		
	Mean	50	26.7	72.8	186	HSDG
272X0	SD	10	8.9	15.0	3	NHSDG
	N	191	191	172		
	Mean	50	27.5	79.3	136	HSDG
324X0	SD	10	10.4	13.4	0	NHSDG
	N	138	138	126		
	Mean	50	34.8	80.1	97	HSDG
328X0	SD	10	15.3	12.5	1	NHSDG
	N	98	94	87		
	Mean	50	28.4	58.6	253	HSDG
423X5	SD	10	10.1	16.4	3	NHSDG
	N	261	261	219		
	Mean	50	31.1	56.1	221	HSDG
426X2	SD	10	12.0	18.9	15	NHSDG
	N	255	239	201		
	Mean	50	22.9	57.0	146	HSDG
492X1	SD	10	12.8	18.8	2	NHSDG
	N	156	156	127		
	Mean	50	28.0	58.3	193	HSDG
732X0	SD	10	11.5	17.4	1	NHSDG
	N	197	197	179		

[a]Sample sizes (N) for each variable of an AFS may not be equal due to missing or invalid data.
[b]Calculated as standard T-Scores.
[c]Calculated as Time in Service (Months).
[d]Calculated as Armed Forces Qualification Test (AFQT) Percentile.
[e]Reported as High School Diploma Graduate (HSDG) or Non-High School Diploma Graduate (NHSDG).

The second index, test-retest reliability, reflects the stability or dependability of the hands-on test scores over time (i.e., from one occasion to another). The result of test-retest analysis for AFS 426X2 is based on data collected during pretesting and full-scale administration of the HOPT. The results reported here are consistent with generally accepted levels for test-retest reliability. The final reliability measure, coefficient alpha, is an estimate of internal consistency. This estimate provides an indication of the extent to which items (tasks) comprising the test are measuring the same concept (i.e., job proficiency). These coefficients are reported along with the associated standard errors of measurement.

Table 2. **Mean HOPT Scores (Standardized T-Scores) by Aptitude and Job Experience for Eight Specialties[a]**

Job Experience (Months)		AFS 122X0		AFS 272X0		AFS 324X0		AFS 328X0		AFS 423X5		AFS 426X2		AFS 492X1		AFS 732X0	
		Upper Half AFQT (I-IIIa)	Lower Half AFQT (IIIb-IV)	Upper Half AFQT (I-IIIa)	Lower Half AFQT (IIIb-IV)	Upper Half AFQT (I-IIIa)	Lower Half AFQT (IIIb-IV)	Upper Half AFQT (I-IIIa)	Lower Half AFQT (IIIb-IV)	Upper Half AFQT (I-IIIa)	Lower Half AFQT (IIIb-IV)	Upper Half AFQT (I-IIIa)	Lower Half AFQT (IIIb-IV)	Upper Half AFQT (I-IIIa)	Lower Half AFQT (IIIb-IV)	Upper Half AFQT (I-IIIa)	Lower Half AFQT (IIIb-IV)
1-12	Mean	41.4	42.3	--	--	--	--	--	--	45.2	44.4	41.9	--	43.3	40.2	42.1	39.3
	SD	16.6	9.4	--	--	--	--	--	--	7.9	--	--	--	11.3	8.5	7.5	6.7
	N	8	13	--	--	--	--	--	--	5	1	1	--	13	14	11	4
13-24	Mean	48.5	47.5	48.2	45.7	47.2	40.5	46.5	34.6	47.8	47.9	48.8	47.0	53.5	47.3	47.5	43.9
	SD	9.2	5.2	10.8	15.8	10.7	--	8.3	--	9.2	7.6	11.2	9.0	6.4	12.2	9.0	9.0
	N	22	16	76	4	59	1	23	1	54	30	48	29	31	20	35	22
25-36	Mean	52.5	50.4	52 6	52.5	49.4	49.0	51.9	--	50.5	48.0	51.8	50.7	56.9	56.1	54.3	49.5
	SD	9.3	9.4	8.2	12.1	9.1	6.1	10.1	--	11.8	10.0	11.9	7.7	8.6	8.8	10.1	7.6
	N	41	20	57	4	34	3	15	--	54	23	31	27	13	7	45	20
37+	Mean	50.8	56.7	52 9	51.0	55.1	--	51.9	--	56.3	49.1	50.7	51.9	53.4	49.8	57.1	49.0
	SD	10.5	6.2	8.1	7.2	7.9	--	9.5	--	9.2	10.1	10.9	8.2	11.2	5.8	8.4	10.4
	N	43	9	28	3	29	--	44	--	33	19	36	29	17	12	25	17
TOTAL GROUP	Mean	50.3	48.8	50.6	49.6	49.7	46.9	50.4	34.6	50.6	48.2	50.1	49.9	52.3	47.2	51.7	46.8
	SD	10.6	9.0	9.7	11.8	10.1	6.6	9.5	--	10.6	8.9	11.2	8.5	9.8	10.7	10.3	9.2
	N	114	58	161	11	122	4	82	1	146	73	116	85	74	53	116	63

[a]Total sample sizes (N) may not correspond to those in Table 1 due to missing or invalid data.

Table 3. **Reliability Estimates and Standard Errors of Measurement for Job-Specific Hands-On Performance Tests**[a]

AFS	Scorer[b]	Test Retest	Coefficient Alpha	SEM[c]
122X0	.92	–	.81	4.36
	8		195	
272X0	.93	–	.74	5.10
	18		191	
324X0	.97	–	.67	5.74
	29		138	
328X0	.97	–	.80	4.47
	20		98	
423X5	.97	–	.65	5.92
	14		261	
426X2	–	.77	.76	4.90
		30	255	
492X1	.94	–	.78	4.69
	20		156	
732X0	.98	–	.67	5.74
	17		197	

[a]Sample size (N) is indicated below each reliability estimate.
[b]Scorer agreement at the step-level.
[c]Standard Error of Measurement calculated using Coefficient Alpha reliability estimates and HOPT standard T-Score standard deviations.

Hands-on Performance Test, Aptitude, Job Experience, and Educational Attainment Relationships. Intercorrelations among these measures for each AFS are presented in Table 4. In general, the better performers were those who had been in the Service longer. The relationship of AFQT and HOPT scores varied by specialty. Table 5 presents the correlations corrected for range restriction in AFQT scores due to selection. A multivariate correction procedure (Mifflin & Verna, 1977) recommended by the National Academy of Sciences and agreed upon by the Joint-Service Job Performance Measurement Working Group was used for the purpose of these analyses.

Air Force-Specific Data Analyses

More extensive data analyses were performed on the first four AFSs to explore further the reliability and validity of the Air Force job performance measures.

Generalizability Theory Analyses. The reliability of the job performance measures was assessed using Generalizability Theory. Two sets of Generalizability Theory analyses were performed to investigate the generalizability of different rating forms, rating sources, and WTPT components. These analyses provide general information about the reliability of the different measures but, more importantly, also provide specific information that reflects the way in which the Air Force could use the measurement instruments.

Table 4. **Intercorrelations Among Measures for Eight Specialties (Sample Value)**

	AFS 122X0 Measure				AFS 272X0 Measure			
Measure	1	2	3	4	1	2	3	4
1. HOPT	1.00	.33	.12	.04	1.00	.25	.10	-.06
2. Job Experience	.33	1.00	.29	.06	.25	1.00	-.05	-.06
3. Aptitude	.12	.29	1.00	.06	.10	-.05	1.00	-.05
4. Educational Attainment	.04	.06	.06	1.00	-.06	-.06	-.05	1.00

	AFS 324X0 Measure				AFS 328X0 Measure			
Measure	1	2	3	4	1	2	3	4
1. HOPT	1.00	.36	.28	.00	1.00	.29	.32	.05
2. Job Experience	.36	1.00	.20	.00	.29	1.00	.00	.06
3. Aptitude	.28	.20	1.00	.00	.32	.00	1.00	.11
4. Educational Attainment	.00	.00	.00	1.00	.05	.06	.11	1.00

	AFS 423X5 Measure				AFS 426X2 Measure			
Measure	1	2	3	4	1	2	3	4
1. HOPT	1.00	.23	.17	-.06	1.00	.18	.10	.10
2. Job Experience	.23	1.00	.07	-.03	.18	1.00	-.14	-.15
3. Aptitude	.17	.07	1.00	-.03	.10	-.14	1.00	.01
4. Educational Attainment	-.06	-.03	-.03	1.00	.10	-.15	.01	1.00

	AFS 492X1 Measure				AFS 732X0 Measure			
Measure	1	2	3	4	1	2	3	4
1. HOPT	1.00	.35	.32	-.03	1.00	.38	.29	-.08
2. Job Experience	.35	1.00	.11	-.12	.38	1.00	.02	-.03
3. Aptitude	.32	.11	1.00	.05	.29	.02	1.00	-.18
4. Educational Attainment	-.03	-.12	.05	1.00	-.08	-.03	-.18	1.00

Table 5. **Intercorrelations Between Measures for Eight Specialties (Corrected for Range Restriction)**

	AFS 122X0 Measure				AFS 272X0 Measure			
Measure	1	2	3	4	1	2	3	4
1. HOPT	1.00	.35	.21	.03	1.00	.20	.16	-.07
2. Job Experience	.35	1.00	.36	.05	.20	1.00	-.23	-.03
3. Aptitude	.21	.36	1.00	-.04	.16	-.23	1.00	-.12
4. Educational Attainment	.03	.05	-.04	1.00	-.07	-.03	-.12	1.00

	AFS 324X0 Measure				AFS 328X0 Measure			
Measure	1	2	3	4	1	2	3	4
1. HOPT	1.00	.59	.66	.00	1.00	.22	.67	.23
2. Job Experience	.59	1.00	.51	.00	.22	1.00	-.09	-.10
3. Aptitude	.66	.51	1.00	.00	.67	-.09	1.00	.34
4. Educational Attainment	.00	.00	.00	1.00	.23	-.10	.34	1.00

	AFS 423X5 Measure				AFS 426X2 Measure			
Measure	1	2	3	4	1	2	3	4
1. HOPT	1.00	.24	.36	-.13	1.00	.02	.29	.11
2. Job Experience	.24	1.00	.09	-.04	.02	1.00	-.41	-.16
3. Aptitude	.36	.09	1.00	-.24	.29	-.41	1.00	.04
4. Educational Attainment	-.13	-.04	-.24	1.00	.11	-.16	.04	1.00

	AFS 492X1 Measure				AFS 732X0 Measure			
Measure	1	2	3	4	1	2	3	4
1. HOPT	1.00	.33	.35	.01	1.00	.32	.53	-.21
2. Job Experience	.33	1.00	-.02	-.15	.32	1.00	.01	-.03
3. Aptitude	.35	-.02	1.00	.29	.53	.01	1.00	-.34
4. Educational Attainment	.01	-.15	.29	1.00	-.21	-.03	-.34	1.00

The first set of analyses investigated the generalizability of information obtained from different rating forms (task, dimensional, global, and Air Force-wide) and different rating sources (incumbents, peers, and supervisors). Results were consistent across the four AFSs. This finding indicated that ratees (incumbents) were rank-ordered similarly across rating forms.

One of the largest sources of variation in the obtained ratings was the interactive relationship between incumbents (ratees) and rating sources (incumbents, peers, and supervisors). This indicates that each source had a unique perspective, in that they tended to rank an incumbent's performance differently from one another. In addition, a measurement condition was simulated that included a single rating source and a single rating form, as is typical of most rating situations and those likely to occur in the Air Force. Under this condition, measurement error was larger, and variance due to true individual differences in performance was smaller relative to the conditions in the Air Force JPM project (three rating sources and four rating forms). This outcome suggests that more reliable data can be collected from all three rating sources.

The second set of analyses investigated the generalizability of WTPT components. Specific areas of interest were methods (hands-on and interview), number of tasks within methods, and number of steps within tasks. Results were again consistent across AFSs. One notable finding is that incumbents were ranked similarly on both the hands-on and interview measures. However, the use of both methods together produced substantially higher reliabilities than did either method used alone, suggesting that both methods be used.

Additional analyses were conducted to approximate the most typical manner in which the WTPT has been used. WTPT includes the use of two methods (hands-on and interview) and 10 tasks for each method, with each task comprised of 15 steps. Under these measurement conditions, the generalizability coefficients were extremely high, with individual differences accounting for 80% to 91% of the variance. This indicates that the WTPT is reliable under the measurement conditions typically used by the Air Force. Further, results indicated that increasing the number of methods, tasks, or steps would not improve reliability substantially and, therefore, would not be cost-effective.

Validity. An exploratory factor analysis procedure was used to determine the underlying structure (construct validity) of the JPMS. Results of these analyses were consistent across the first four AFSs studied. The five factors identified were technical proficiency, interpersonal proficiency, supervisor ratings, peer ratings, and self-ratings. The correlational analyses presented in Table 6 indicate the predictability of the AFQT for these factors. The AFQT predicted technical proficiency for AFS 328X0 and technical proficiency, supervisor ratings, and peer ratings for AFS 492X1.

Table 6. **Correlations Between AFQT and JPMS Factors for Four AFSs (Sample Value)**

JPMS Factors	AFS			
	272X0	328X0	426X2	492X1
Technical proficiency	.11	.33	.18	.37
Interpersonal proficiency	-.04	.14	.08	.19
Supervisor ratings	.04	.18	.04	.28
Self-ratings	-.09	-.02	.04	.16
Peer ratings	.04	.14	.11	.36

IV. PLANS AND DIRECTIONS

The Air Force's ongoing and planned R&D in the Job Performance Measurement area can be described in terms of four research thrusts: advanced development of the Job Performance Measurement technology, specification of an operational JPMS, application of JPM technologies to other areas, and development of performance-based selection and classification models.

Considerable work remains in understanding the reliability, validity, and utility of the Job Performance Measurement technology developed by the Air Force. In general, the key research questions center around assessing the quality of the individual performance measures. Research plans for the next few years include expanded application of Generalizability Theory to all eight AFSs. An assessment of the costs and benefits of the measurement system components is planned, with benefits defined as the reliability and validity, interrelationships, practicality, and acceptability of the measures. Research is planned for developing technologies to interpret performance data in terms of acceptable levels of performance (i.e., minimal competence). Finally, it appears that an ability to translate the performance scores into other metrics, such as dollars or units of productivity, is important to understanding the utility of performance data. Thus, research will be conducted in this area as well.

A wide range of studies are needed before the JPM technology can be used by the Air Force to collect job performance data routinely and in a cost-effective manner. In addition to the above work, we plan to initiate a review of the methods currently used throughout the Air Force for recording individual job performance. Future efforts will examine the quality of these measures and compare them with those generated by the JPMS. From this and previous studies will come the guidelines on the measurement techniques to use in gathering performance information for given purposes (e.g., enlistment standard setting or training feedback). This line of research will also focus on the mechanisms for collecting the performance data. Existing and planned automated systems, such as those for maintaining personnel records, may contain performance information that could be routinely accessed. Such systems must be identified and evaluated. However, where they do not exist, procedures will be outlined for collecting and maintaining the needed performance information.

The third research thrust examines how performance information might be integrated into the Air Force training system as a source of feedback for identifying training needs and evaluating training programs. Given that the goal of our technical training programs is to prepare individuals to be capable of performing their duties, job performance forms the most reasonable criterion for evaluating how well training has met this goal. Research in this area will examine existing methods for identifying training needs which help define training course content, and determine how job performance could be used to clarify these needs by identifying areas of overtraining and undertraining. It will also focus on the types of performance information that would be needed for this purpose, because such information will likely differ in specificity, amount, and level of detail from that necessary for establishing selection and classification standards. Application of the JPM technology for training program evaluation has begun; the performance of graduates of an experimental course providing additional training for Jet Engine Mechanics is being evaluated with hands-on and knowledge tests. The results of this study will help training managers decide whether to continue the course and expand the program to other specialties. In addition, JPMS hands-on, interview, and job knowledge test development procedures have been successfully used to develop evaluation instruments for task certification within another R&D project designing an Advanced On-the-job Training System for the Air Force.

The last area deals with the Air Force's research plans for examining the value of integrating job performance information into the selection and classification system. As with instituting a JPMS, a considerable amount of work must be done before a performance-based system for setting accession standards can be developed. Initial research will use the data collected on

the eight AFSs and focus on whether ASVAB prediction improves as a result of using job performance information instead of or in addition to currently used technical training scores. Specific analyses will examine the relationship between performance and ASVAB when different combinations of the ASVAB subtests are used. Optimal performance-based ASVAB composites will be compared to those being used to predict training success. In addition, new predictor research will include an examination of the extent to which scores on the new tests relate to job performance. This stream of research will aid in deciding the value of developing a JPMS with selection and classification applications in mind or whether the present system is efficient enough given our technologies. Finally, our plans are to evaluate the outcomes of Air Force classification models currently under development and testing (e.g., the Processing and Classification of Enlistees model) when job-performance-based information is included, to determine the value added.

V. SUMMARY AND CONCLUSIONS

The Air Force research efforts in the Job Performance Measurement/Enlistment Standards Project have resulted in the development of a state-of-the-art technology for assessing the performance capability of airmen which incorporates several different appraisal methods. The highest fidelity measure is a hands-on work sample test, against which the other measures (performance interviews, rating forms, and job knowledge tests) are being evaluated. Analyses are underway on data collected on eight Air Force enlisted specialties. Results indicate that the ASVAB scores do relate to individual job performance measures. Although the relationship between AFQT scores and hands-on performance varies from one specialty to another, each of the correlations found within the eight AFSs studied has been positive, ranging from .16 to .67 after correction for restriction in range. Analyses of the data have revealed that the Walk-Through Performance Test is a reliable method for measuring the technical proficiency of airmen. Studies should continue to examine the relationships among the various methods and the relative contributions to the measurement of job performance. Air Force research plans include determining how to collect job performance information more effectively and how to use this information in setting selection and classification standards, and as training feedback.

REFERENCES

Hedge, J.W., & Teachout, M.S. (1986, November). *Job performance measurement: A systematic program of research and development* (AFHRL-TP-86-37, AD-A174 175). Brooks AFB, TX: Training Systems Division, Air Force Human Resources Laboratory.

Lipscomb, M.S., & Hedge, J.W. (1988, June). *Job performance measurement: Topics in the performance measurement of enlisted personnel* (AFHRL-TP-87-58, AD-A195 630). Brooks AFB, TX: Training Systems Division, Air Force Human Resources Laboratory.

Mifflin, T.L., & Verna, S.M. (1977). *A method to correct correlation coefficients for the effects of multiple curtailment* (CRC-336). Arlington, VA: Marine Corps Operations Analysis Group, Center for Naval Analyses.

* U. S. GOVERNMENT PRINTING OFFICE: 1991--761-052 40006